

College of Engineering



Drexel E-Repository and Archive (iDEA)

<http://idea.library.drexel.edu/>

Drexel University Libraries

www.library.drexel.edu

The following item is made available as a courtesy to scholars by the author(s) and Drexel University Library and may contain materials and content, including computer code and tags, artwork, text, graphics, images, and illustrations (Material) which may be protected by copyright law. Unless otherwise noted, the Material is made available for non profit and educational purposes, such as research, teaching and private study. For these limited purposes, you may reproduce (print, download or make copies) the Material without prior permission. All copies must include any copyright notice originally included with the Material. **You must seek permission from the authors or copyright owners for all uses that are not allowed by fair use and other provisions of the U.S. Copyright Law.** The responsibility for making an independent legal assessment and securing any necessary permission rests with persons desiring to reproduce or use the Material.

Please direct questions to archives@drexel.edu

Bandwidth Constraint Models: A performance study with preemption on link failures

Joshua B. Goldberg, Sukrit Dasgupta, and Jaudelice C. de Oliveira

Electrical and Computer Engineering Department

Drexel University

Philadelphia, PA 19104

Email: {jbg25,sd88,jco24}@drexel.edu

Abstract—Bandwidth constraint models have been a topic of intense discussions at the IETF meetings. Three conventional methods have been described in informational IETF RFCs and their performance on a single link has been analyzed and discussed in the literature. In this article, we take a further step into analyzing their performance and optimal bandwidth constraint setting for a real network scenario. A new model is proposed and compared to existing ones when failure events may cause preemption of traffic trunks in a network. Our simulations results provide great insight on the benefits of the methods.

I. INTRODUCTION AND RELATED WORK

The Internet has been evolving from an all best-effort service network to a more sophisticated network in which differentiated service provisioning with multiple levels of quality of service (QoS) is a must. DiffServ-aware Multiprotocol Label Switching (MPLS) Traffic Engineering (DS-TE) is a natural solution for deploying traffic engineering (TE), and it is capable of handling most of the recent issues raised in IP-based networks. DS-TE enforces different Bandwidth Constraints (BCs) on different classes, rather than on an aggregate basis across all classes, to achieve per-class traffic engineering.

RFC 4124 [1] specifies the IGP and RSVP-TE signaling extensions for support of DS-TE requirements given in [2]. While [2] provides the requirements and selection criteria for BCs Models for the allocating bandwidth to individual classes, [1] does not specify nor assume a particular BC Model. However, the extensions for DS-TE specified in RFC 4124 do support the Russian Dolls Model (RDM) specified in [3], the Maximum Allocation Model (MAM) specified in [4], and the Maximum Allocation with Reservation Model (MAR) specified in [5].

While engaging discussions were held at the IETF Traffic Engineering working group meetings about which of the models should be required for deployment, only preliminary work had been carried out to compare the performance of the methods. In [6], the author provides a performance comparison between RDM and MAM on a single link using isolation across class-types as the metric of interest. While it was pointed out that some network providers may decide not to deploy preemption, the recent increased interest in preemption policies stress their importance, and therefore justifies the use of a BC model that works well when full preemption is enabled.

Preemption policies have gained renewed attention in recent years as a flexible and effective control mechanism to dynamically allocate capacity among competing traffic classes with different priorities. In an MPLS context, a *preemption attribute* determines whether a Label Switched Path (LSP) with a certain *priority attribute* can preempt another LSP with a lower *priority attribute* from a given path when there is a competition for available resources. The preempted LSP may then be rerouted.

In 1992, Garay and Gopal [7] addressed the call preemption problem in communication networks. The authors showed that the problem of selecting a connection/trunk for preemption in order to minimize the number of preempted connections or minimizing the amount of bandwidth preempted is NP-complete. They proposed heuristics for a centralized network framework. Citing Garay and Gopal's work, Peyravian and Kshemkalyani [8] proposed decentralized network connection preemption algorithms which optimize three fixed criteria in a given order of importance: number of connections, bandwidth, and priority; and bandwidth, priority, and number of connections. These decentralized policies were the basis for the author's work on flexible and adaptive preemption policies [9], in which an order of importance for the considered criteria is not fixed, but can be configured by the network provider according to the network's best interest. In [10], the authors presented an algorithm concentrating on bandwidth allocation and management with preemption. A BC model similar to MAR [5], is proposed and implemented using three colored matrices. The algorithm is however centralized (the matrices need to be advertised). Preemption is performed until a connection reaches its minimum bandwidth, in which case preemption is not allowed. Stanisic and Devetsikiotis [11] proposed simple preemption policies based on random selection, which reduce the time needed to select a set of connections to be preempted, a very interesting feature for large topologies. [12] focuses on a routing algorithm which considers preemption mechanisms. A path is selected based on the number of connections (or LSPs in this case) which need to be preempted. The routing algorithm therefore tries to minimize the preemption events and therefore the need for rerouting.

While several researchers have concentrated on proposing new preemption policies, little attention has been given to

the bandwidth constraint model problem. [6] was a first step into evaluating the performance of RDM and MAM, and [10] proposed a centralized BC model similar to MAR. In this paper, we extend the analysis in [6] to a real network case with full preemption deployment, and investigate several other performance metrics of interest. Besides RDM and MAM, MAR and a newly proposed decentralized BC model, called Blocking Constraint Model (BCM), are also studied and compared to a case in which full sharing of the bandwidth is allowed (a model we refer to as NULL). The BCM is similar to MAM but introduces new preemption permissions on the last hop to the destination, as a path computation failure on the last hop reduces the chances of other routes being found. The models are carefully configured so as to give the same protection to each priority level, as accurately as possible. We investigate the number of preempted and blocked LSPs, as well as the number of LSPs not on their shortest path as a comparison metric for all the BC models and for the absence of a model (NULL).

The rest of the paper is organized as follows. In Section II, the RDM, MAM, and MAR BC models are introduced along with the proposed BCM and also NULL. The performance metrics are discussed in Section III. In Section IV, the simulation scenario is described in detail and the results are presented and discussed in Section V. Section VI concludes the paper.

II. BANDWIDTH CONSTRAINT MODELS

The fundamental requirement for DS-TE is to be able to enforce different bandwidth constraints for different sets of traffic classes. In [2], a Class-Type (CT) is defined as the set of traffic trunks crossing a link in which a specific set of bandwidth constraints is enforced. DS-TE must allow support for up to 8 CTs: CT_c , $c = 0, \dots, 7$. By definition, each CT is assigned either a Bandwidth Constraint (BC), or a set of BCs. Therefore, DS-TE must support up to 8 BCs: BC_b , $b = 0, \dots, 7$. However, a smaller number of BCs may be activated in a network. We will explain the BC models in the context of MPLS, and therefore will make reference to LSPs being assigned CTs.

A. Russian Doll Model (RDM)

RDM may be defined as follows.

- Maximum number of BCs is equal to maximum number of CTs = 8;
- All LSPs from CT_c must use no more than BC_b (with $b \leq c \leq 7$, and $BC_b \leq BC_{b-1}$, for $b = 1, \dots, 7$), i.e.:
 - All LSPs from CT_7 use no more than BC_7 ;
 - All LSPs from CT_6 and CT_7 use no more than BC_6 ;
 - All LSPs from CT_5 , CT_6 and CT_7 use no more than BC_5 ;
 - ...
 - All LSPs from CT_0 , CT_1 , CT_2 , CT_3 , CT_4 , CT_5 , CT_6 and CT_7 use no more than BC_0 .

RDM is a good match to the way many network operators manage QoS in the data plane, e.g., voice in Low Latency

Queuing (LLQ), business data in a high weight Class-Based Weighted Fair Queueing (CBWFQ) class, and best effort getting whatever is left. RDM provides for good isolation between classes, and efficient use of bandwidth.

B. Maximum Allocation Model (MAM)

MAM simply gives an upper limit on the amount of bandwidth available to each class of traffic (CT_i uses no more than BC_i , $i = 0, \dots, 7$). In MAM, each bandwidth pool can have its own overbooking ratio. However, unused bandwidth in one pool is not available for other pools.

C. Maximum Allocation with Reservation Model (MAR)

MAR gives a soft upper limit on the amount of bandwidth available to each class of traffic. This limit is soft because it can be disregarded if the link is not above a certain usage r . This level of usage is defined by the reservation for the link ($r = LinkCapacity - Reservation$). As long as the link usage is below r , any LSP is allowed on the link. Once usage grows above this value, MAR behaves exactly in the same manner as MAM.

D. Blocking Constraint Model (BCM)

We propose a new model called *Blocking Constraint Model* (BCM). BCM works similarly to MAM, except it defines three new classes of traffic. A new class of traffic is defined for each priority. An LSP can be allowed into the new class if the link in consideration is the last hop of the LSP to the destination (last link). The idea is to give special rights to LSPs on these links since a denial on the last link of the path eliminates a lot of possible paths. With BCM, preemption is performed in the following order:

- High Priority (HP) can preempt Low Priority (LP) then Medium Priority (MP)
- MP can preempt LP
- LP cannot preempt
- HP on last link (HP-LL) can preempt HP, MP, LP, Medium Priority Last Link (MP-LL), Low Priority Last Link (LP-LL)
- MP-LL can preempt HP, MP, LP, LP-LL
- LP-LL can preempt HP, MP, LP

In comparison, BCM works exactly like MAM on any link except the last link on a path. On the last link that connects an LSP to the destination, BCM upgrades the preemption permissions of the LSP.

E. No Bandwidth Constraint Model (NULL)

NULL is the case where no BC model is used. In this case, default preemption is used and any LSP can use any portion of the link (CT_i uses no more than $BC_i = 100\%$, $i = 0, \dots, 7$). Bandwidth is completely shared.

III. PERFORMANCE METRICS

Since TE-LSPs are associated with reserved bandwidth, their routes are also affected by the configuration of the BC model in place. The following metrics are studied for performance comparisons on a per priority basis.

Number of Preempted TE-LSPs: This metric captures the total number of TE-LSPs that have undergone preemption with respect to time. The absolute number is kept track of and may include the same TE-LSP being accounted for more than once since it underwent preemption at different times.

Number of blocked TE-LSPs per priority: This time varying metric captures the number of TE-LSPs that failed to find a route after being preempted by higher priority TE-LSPs.

Number of TE-LSPs not on the shortest path: This is a per priority analysis of how the number of TE-LSPs not on the shortest path varies with time and how it is dependent on the BC model used. Let $\mathcal{P} = \{p_1^t, p_2^t, \dots, p_n^t\}$ form the set of instantaneous paths of all the TE-LSPs. Let the identity function $\mathbf{I}(p_i^t) = 1$ if the path p_i^t of TE-LSP i at instant t is its shortest path for its corresponding reserved bandwidth. This metric captures the behavior of $\sum_{i=0}^n \mathbf{I}(p_i^t)$ with time.

IV. SIMULATION SETUP

A very detailed event based simulator has been developed to replicate a real life network scenario accurately. The BC models govern the behavior of the entities and have been carefully configured to provide the same platform for comparison. Following is the set of entities used in the simulation with a brief description of their characteristics.

A. Simulator Entities and Events

Node: Every node in the simulator has a knowledge of the topology in the form of an adjacency matrix. The routing functionality on the node uses the topology information to compute routes using CSPF. Every node also maintains state for every TE-LSP passing through it.

Link: Every link is associated with two nodes, one on each of its ends. A link maintains information for the amount of utilization and reservation arising out of all the TE-LSPs that traverse it.

TE-LSP: A TE-LSP is characterized by its reserved size, and a source and destination node. A route is assigned to it by the routing functionality present in a source node. A TE-LSP has a priority associated with it that has been assigned to it by virtue of its size. Tables I describes the network topology and the distribution of TE-LSP sizes and priorities, respectively. The network topology has been obtained from Rocketfuel.

Failure and Restore Events: A failure event represents the failure of a link or a node in the network. When a link fails, TE-LSPs traversing it get rerouted. Reservations held by the TE-LSP are first removed from the old route before they are setup on a new route. A restore event represents a failed link getting restored. For the simulations, a failure event takes place every 10 units of time and a restore event takes place 5 units of time after the corresponding failure event. Only one link is down at any point of time in the network.

Routing and Blocking Events: A routing event initiates the routing of a TE-LSP. This can arise from a TE-LSP being setup initially at the beginning of the simulation, during a link failure or due to a preemption. A blocking event takes place when a TE-LSP fails to find a route.

Preemption Event: A preemption event occurs when a higher priority TE-LSP causes one or more TE-LSPs of lower priorities to get preempted by a midpoint router on the route that the higher priority TE-LSP is trying to be setup on.

TABLE I
TOPOLOGY AND TE-LSP DETAILS USED IN SIMULATIONS

Network Description					TE-LSP Description			
# of nodes	# of links	OC3 links	OC48 links	OC192 links	Number	100Kb-1Mb	1-20Mb	20-50Mb
83	167	0	132	35	6806	70%	25%	5%
					Priority	2	1	0

B. Configuration of Bandwidth Constraint Models

The BC models have been carefully configured so as to give, as much as possible, the same protection to each priority level. With the configuration described below, in a highly congested link HP traffic would have 40% of the link, while MP would have 50%, and low priority would have 10% after preemption has taken place.

- **RDM:** $BC_2 = 40\%$ (HP), $BC_1 = 90\%$ (HP+MP), $BC_0 = 100\%$ (HP+MP+LP).
- **MAM:** $BC_2 = 40\%$ (HP), $BC_1 = 90\%$ (MP), $BC_0 = 100\%$ (LP).
- **MAR:** $BC_2 = 40\%$ (HP), $BC_1 = 90\%$ (MP), $BC_0 = 100\%$ (LP), Reservation=90%.
- **BCM:** $BC_2 = 40\%$ (HP), $BC_1 = 90\%$ (MP), $BC_0 = 100\%$ (LP), $P'_h = 60\%$, $P'_m = 90\%$, $P'_l = 100\%$.

V. SIMULATION RESULTS

A. Number of Preempted TE-LSPs

Fig. 1, shows that when no BC model is employed (NULL), the largest number of low priority preemptions take place. This is because low priority TE-LSPs are unprotected from getting preempted by higher priority TE-LSPs. In MAR no constraint is held on the link before it reaches 10% utilization. This leads to 90% of the link being controlled by the model. In the configuration of the BC model, high priority is constrained to 40% and medium priority is constrained to 90%. When a link is about to be fully utilized, the unprotected 10% becomes available to medium or high priority TE-LSPs, conditioned to their respective constraints. If the unprotected 10% consists of low priority TE-LSPs, they are preempted by any incoming higher priority TE-LSPs. This results in the medium priority TE-LSPs being protected from the high priority TE-LSPs which is close to the behavior of no BC model in place. For the low priority TE-LSPs, compared to NULL, MAR offers more protection from high priority TE-LSPs and similar protection from medium priority TE-LSPs. In MAM and RDM, high priority TE-LSPs are constrained to 40% of the link capacity thus guaranteeing 60% of the link capacity to medium priority TE-LSPs for MAM and 50% of the link capacity for RDM. The small difference explains the number of medium priority (Fig. 2) preemptions being noticeably lower on an average for MAM than for RDM. In RDM, 10% of the link capacity is always reserved for low priority traffic, which results in protection of a large number of low priority TE-LSPs (as they

are very small in size) and hence there are significantly less low priority preemptions. In comparison, as configured, MAM offers no such protection thereby resulting in more low priority preemptions. To reduce blocking, BCM allows TE-LSPs with priority P'_m and P'_l to preempt TE-LSPs with priority P_m and P_l . This leads to more medium priority preemptions in BCM when compared to MAM as seen in Fig 2. As shown in Fig. 1, BCM also results in fewer lower priority preemptions when compared to MAM and MAR.

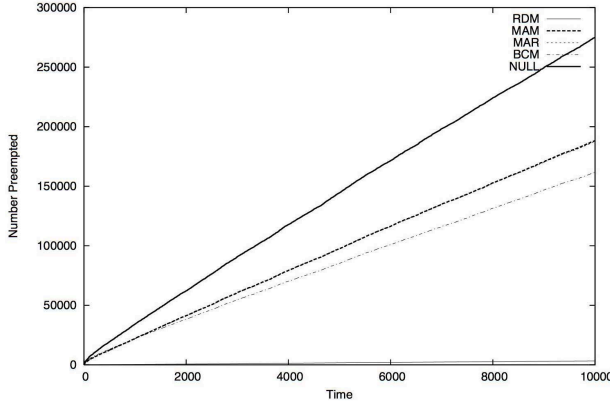


Fig. 1. Total number of low priority TE-LSPs preempted

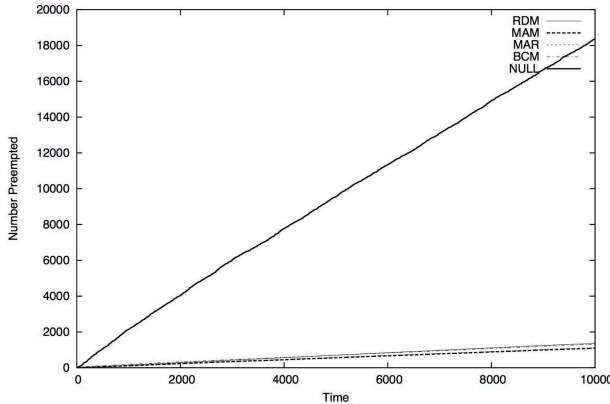


Fig. 2. Total number of medium priority TE-LSPs preempted

B. Number of blocked TE-LSPs per priority

As can be seen from Fig. 3, when no BC model is employed (NULL), the largest number of low priority TE-LSPs are blocked since they are unprotected. The small difference between RDM and MAM (10% extra room allowed for MP in MAM) explains the number of blocked medium priority TE-LSPs being negligible on an average for RDM and none for MAM. In RDM, 10% of the link capacity is always reserved for low priority traffic. In comparison, MAM offers no such protection thereby resulting in more blocking for low priority TE-LSPs. To reduce blocking, BCM allows TE-LSPs with priority P'_m and P'_l to preempt TE-LSPs with priority P_m and P_l . Blocking of low priority TE-LSPs is much less in BCM when compared to MAM as seen in Fig. 3. There may be rare

instances with BCM where a medium priority TE-LSP gets preempted and then blocked. This behavior is seen in Fig. 4.

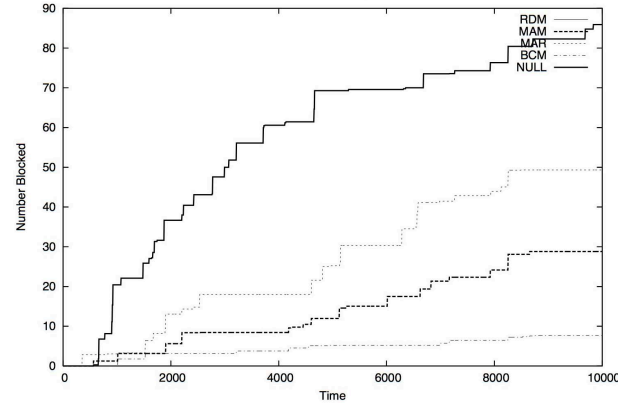


Fig. 3. Total low priority TE-LSPs blocked on preemption

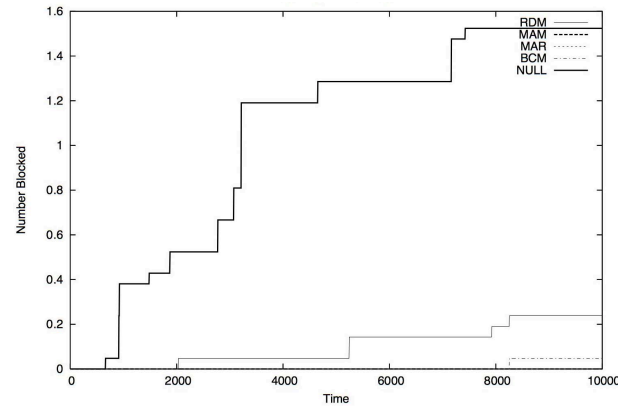


Fig. 4. Total medium priority TE-LSPs blocked on preemption

C. Number of TE-LSPs not on the shortest path

Fig. 5 shows the number of high priority TE-LSPs not on their shortest path. A larger number obviously means more TE-LSPs are on the longer path, which is not favorable. Fig. 5 shows that in the scenario where no BC model is employed, all high priority TE-LSPs try to preempt as many lower priority TE-LSPs as possible in order to get on shorter paths. This results in more TE-LSPs of lower and medium priority not being on their shortest path as shown in Fig. 6 and Fig. 7. BCM, under conditions specified in Section IV, allows P'_h to occupy 60% instead of its usual configured 40% constrained by other priorities. This causes more high priority TE-LSPs to be on their shorter paths than the other BC models. The remaining models, namely MAM, MAR and RDM treat high priority TE-LSPs the same way and show similar behavior for this metric.

MAM and MAR treat medium priority exactly the same way and hence show similar behavior as shown in Fig. 6. BCM causes more medium priority TE-LSPs to follow longer paths since it causes TE-LSPs to move off shorter paths to reduce blocking. RDM only allows 50% occupation of the

link capacity to medium priority TE-LSPs, causing more of them to be on longer paths than BCM. When no constraint model is in place, all medium priority flows can be preempted by high priority TE-LSPs thus causing the largest number of them to be on longer paths.

As shown in Fig. 7, RDM shows the best performance for this metric with respect to low priority TE-LSPs. This is because RDM offers 10% reservation of link capacity for the low priority TE-LSPs. BCM is better than MAM and MAR since in some rare situations, low priority TE-LSPs can preempt medium TE-LSPs to lessen blocking. MAM and MAR offer no protection to low priority TE-LSPs but also do not give high priority TE-LSPs a free reign to preempt them resulting in better performance when no constraints are in place.

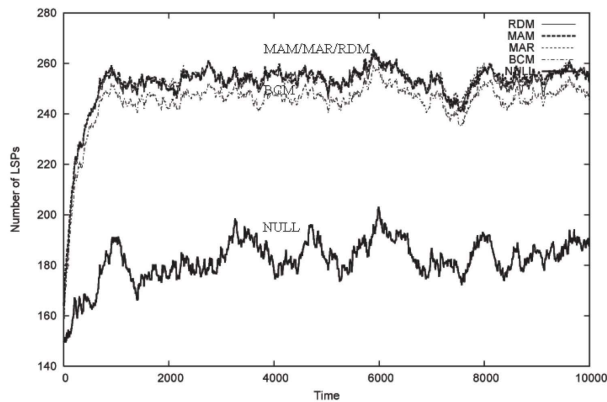


Fig. 5. Number of high priority TE-LSPs not on the shortest path

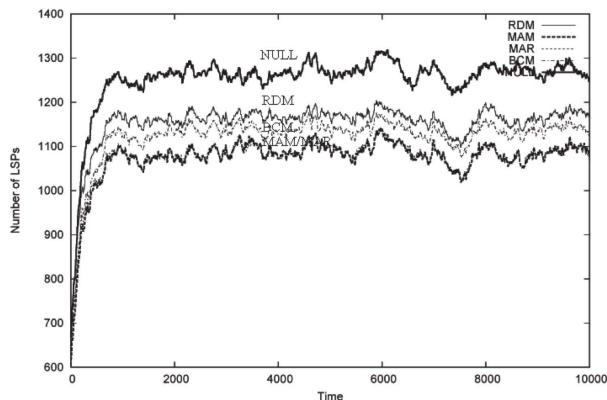


Fig. 6. Number of medium priority TE-LSPs not on the shortest path

VI. CONCLUSION

We studied the behavior of several BC models: RDM, MAM, MAR, a new model BCM, and the absence of a model (NULL) in a real network scenario. The models were carefully configured to give, as much as possible, the same protection level to each Class Type, so that comparison would be possible. We investigated the number of preempted and blocked LSPs, and the number of TE-LSPs not on the shortest

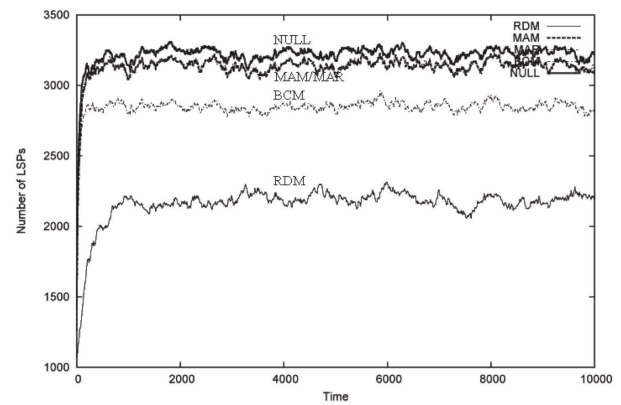


Fig. 7. Number of low priority TE-LSPs not on the shortest path

path for each Class Type. The simulation results show in particular that the RDM model, as configured, had the best performance across the metrics. We proposed a new model, BCM, aimed at reducing the blocking probability of TE-LSPs by allowing preemption across Class-Types on tail-end links. The simulations results provide great insight on the benefits of each model.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers. This work is supported in part by the National Science Foundation under Grant No. 0435247. The opinions and findings in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] F. Le Faucheur, *Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*, IETF, RFC 4124, June 2005.
- [2] F. Le Faucheur, *Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering*, IETF, RFC 3564, July 2003.
- [3] F. Le Faucheur, *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*, IETF, RFC 4127, June 2005.
- [4] F. Le Faucheur and W. Lai, *Maximum Allocation Bandwidth Constraints Model for Diffserv-aware Traffic Engineering*, IETF, RFC 4125, June 2005.
- [5] J. Ash, *Max Allocation with Reservation Bandwidth Constraints Model for DiffServ-aware MPLS Traffic Engineering and Performance Comparisons*, IETF, RFC 4126, June 2005.
- [6] W. Lai, *Bandwidth Constraints Models for Differentiated Services (DiffServ)-aware MPLS Traffic Engineering: Performance Evaluation*, IETF, RFC 4128, June 2005.
- [7] J. A. Garay and I. S. Gopal, "Call Preemption in Communication Networks," in *Proceedings of IEEE INFOCOM*, 1992, pp. 1043–1050.
- [8] A. D. Kshemkalyani M. Peyravian, "Connection Preemption: Issues, Algorithms, and a Simulation Study," in *Proceedings of IEEE INFOCOM*, 1997.
- [9] J. C. de Oliveira, C. Scoglio, I. F. Akyildiz, and G. Uhl, "New Preemption Policies for DiffServ-Aware Traffic Engineering to Minimize Rerouting in MPLS Networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 4, 2004.
- [10] S. Tong, D. Hoang, and O. Yang, "Bandwidth allocation and preemption for supporting differentiated-service-aware traffic engineering in multi-service networks," in *Proceedings of IEEE ICC*, 2002.
- [11] V. Stanicic and M. Devetsikiotis, "A dynamic study of providing quality of service using preemption policies with random selection," in *Proceedings of IEEE ICC*, 2003.
- [12] F. Blanchy, L. Melon, and G. Leduc, "Routing in a MPLS network featuring preemption mechanisms," in *Proceedings of 10th International Conference on telecommunications*, 2003.